PCT ORGANISATION MONDIALE DE LA PROPRIETE INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets 6 : **H04L 12/00**

A2

(11) Numéro de publication internationale: **WO 98/33297**

(43) Date de publication internationale: 30 juillet 1998 (30.07.98)

(21) Numéro de la demande internationale: PCT/FR98/00110

(22) Date de dépôt international: 22 janvier 1998 (22.01.98)

(30) Données relatives à la priorité:
97/00757 24 janvier 1997 (24.01.97) FR

(71) Déposant *(pour tous les Etats désignés sauf US)*: BULL S.A. [FR/FR]; 68, route de Versailles, F–78430 Louveciennes (FR).

(72) Inventeur; et
(75) Inventeur/Déposant *(US seulement)*: PEPING, Jacques [FR/FR]; 72, rue Victor Basch, F–78220 Viroflay (FR).

(74) Mandataire: DENIS, Hervé; Bull S.A., 68, route de Versailles, F–78430 Louveciennes (FR).

(81) Etats désignés: JP, US.

Publiée
Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.

(54) Title: COMPUTER SYSTEM WITH DISTRIBUTED DATA STORING

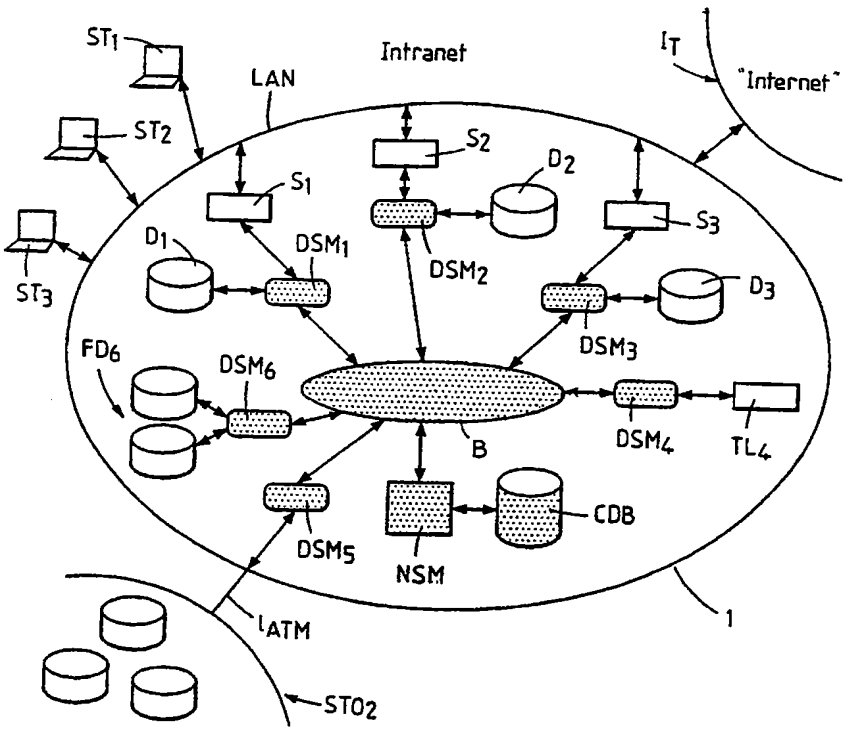
(54) Titre: SYSTEME INFORMATIQUE A STOCKAGE DE DONNEES DISTRIBUE

(57) Abstract

The invention concerns a computer system (1) in which each data storing resource (D_1 to D_3, FD_6, TL_4, STO_e) is under the control of a distributed management unit (DSM_1 to DSM_6), associated or not with a server (S_1 to S_3). The latter receives from a centralised management unit (NSM) control programmes dynamically allocating to it a virtual memory space comprising local storing units and all or part of the external storing resources. The invention is applicable to computer systems.

(57) Abrégé

Dans le système informatique (1) selon l'invention, chaque ressource de stockage de données (D_1 à D_3, FD_6, TL_4, STO_e) est sous la commande d'une unité d'administration distribuée (DSM_1 à DSM_6), associée ou non à un serveur (S_1 à S_3). Celle–ci reçoit d'une unité d'administration centralisée (NSM) des programmes de commande lui attribuant dynamiquement un espace mémoire virtuel comprenant des unités de stockage locales et tout ou partie des ressources de stockage externes.



UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	**ES**	Espagne	**LS**	Lesotho	**SI**	Slovénie
AM	Arménie	**FI**	Finlande	**LT**	Lituanie	**SK**	Slovaquie
AT	Autriche	**FR**	France	**LU**	Luxembourg	**SN**	Sénégal
AU	Australie	**GA**	Gabon	**LV**	Lettonie	**SZ**	Swaziland
AZ	Azerbaïdjan	**GB**	Royaume-Uni	**MC**	Monaco	**TD**	Tchad
BA	Bosnie-Herzégovine	**GE**	Géorgie	**MD**	République de Moldova	**TG**	Togo
BB	Barbade	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tadjikistan
BE	Belgique	**GN**	Guinée	**MK**	Ex-République yougoslave de Macédoine	**TM**	Turkménistan
BF	Burkina Faso	**GR**	Grèce			**TR**	Turquie
BG	Bulgarie	**HU**	Hongrie	**ML**	Mali	**TT**	Trinité-et-Tobago
BJ	Bénin	**IE**	Irlande	**MN**	Mongolie	**UA**	Ukraine
BR	Brésil	**IL**	Israël	**MR**	Mauritanie	**UG**	Ouganda
BY	Bélarus	**IS**	Islande	**MW**	Malawi	**US**	Etats-Unis d'Amérique
CA	Canada	**IT**	Italie	**MX**	Mexique	**UZ**	Ouzbékistan
CF	République centrafricaine	**JP**	Japon	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Pays-Bas	**YU**	Yougoslavie
CH	Suisse	**KG**	Kirghizistan	**NO**	Norvège	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	République populaire démocratique de Corée	**NZ**	Nouvelle-Zélande		
CM	Cameroun			**PL**	Pologne		
CN	Chine	**KR**	République de Corée	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Roumanie		
CZ	République tchèque	**LC**	Sainte-Lucie	**RU**	Fédération de Russie		
DE	Allemagne	**LI**	Liechtenstein	**SD**	Soudan		
DK	Danemark	**LK**	Sri Lanka	**SE**	Suède		
EE	Estonie	**LR**	Libéria	**SG**	Singapour		

Système informatique à stockage de données distribué.

La présente invention concerne un système informatique à stockage de données distribué.

Elle s'applique à un système informatique de traitement de données à architecture réseau et plus particulièrement à une architecture réseau du type dit "INTRANET", desservant une entreprise ou une organisation.

Il est généralement admis qu'un des facteurs clés de la bonne santé d'une entreprise ou d'une société est directement dépendant des informations qu'elle possède. Ce terme "informations" doit être compris dans son sens le plus général. Il s'agit d'informations internes à la société (prix de ventes de produits, gammes de fabrication, etc.) ou de provenance extérieure à celle-ci (données diverses sur la concurrence, de type commercial ou technique, etc.).

Dans l'état des techniques actuel, le traitement de ces informations fait naturellement appel à des systèmes informatiques de plus en plus puissants et complexes. La baisse rapide des prix du matériel, notamment des mémoires de masse (disques durs, unités de bandes magnétiques ou de cartouches, disques optiques) permet de stocker de plus en plus de données, en local ou dans des sites éloignés.

De plus en plus, les systèmes de traitement de données sont fédérés en réseaux. Parmi ceux-ci, on doit signaler le réseau "INTERNET" qui permet le dialogue entre des millions d'ordinateurs disséminés à travers le monde, y compris de simples micro-ordinateurs à usage domestique.

De la même façon, les sociétés utilisent des réseaux locaux spécifiques appelés "INTRANET", qui relient entre elles les différentes ressources informatiques d'un ou plusieurs sitesqui leur sont propres.

De ce fait, elles sont placées devant la nécessité urgente de maîtriser le flot croissant d'informations entrantes et, notamment, de les stocker en des "endroits" où elles peuvent être aisément accessibles, déplacées et administrées de façon la plus efficace possible, et ce au moindre coût.

Ces données doivent également être protégées, dans le sens le plus large de ce terme. On utilise en général le concept connu sous le sigle "D.I.C.", pour Disponibilité, Intégrité, Confidentialité.

Il est en effet nécessaire d'assurer l'intégrité des données, que ce soit contre les défaillances des systèmes ou contre les actions malveillantes. Il est aussi nécessaire de prendre des mesures de "réservation", car certaines données peuvent être confidentielles, du moins à accès limité à des utilisateurs autorisés. Enfin, elles doivent être le plus disponible possible, ce qui implique notamment des mesures de sauvegarde ou une certaine redondance dans le stockage, pour pallier aux déficiences matérielles ou aux erreurs logicielles.

Enfin, il est nécessaire que, une fois le choix d'un système de stockage effectué, la pérennité du système soit assurée. Notamment, on doit pouvoir prendre en compte les technologies futures sans modifications importantes du système.

L'invention se fixe pour but un système visant à satisfaire les besoins évoqués.

Pour ce faire, elle consiste à assurer un stockage distribué des données mettant en oeuvre un adressage virtuel généralisé des ressources de stockage réparties dans un système informatique.

L'invention a donc pour objet un système informatique comprenant une pluralité de moyens de stockage de données distribués et au moins un serveur de stockage desdites données, caractérisé en ce qu'il comprend des moyens d'attribution à chacun desdits serveurs de stockage de données d'un espace mémoire virtuel dont l'étendue est au moins égale à la capacité cumulée de tout ou partie desdits moyens de stockage de données distribués.

Selon une variante de réalisation préférée de l'invention, le système comprend essentiellement des moyens d'administration centralisés et des moyens d'administration distribués associés à au moins une ressource de stockage du système informatique, ainsi qu'un bus de communication à grande vitesse reliant ces moyens d'administration distribués, entre eux, et avec les moyens centralisés.

L'invention sera mieux comprise et d'autres caractéristiques et avantages apparaîtront à la lecture de la description qui suit en référence aux figures annexées, parmi lesquelles :

- La figure 1 illustre schématiquement un exemple d'architecture de traitement de donnée de type réseau selon l'art connu ;

- la figure 2 est un chronogramme illustrant l'échange de données entre deux serveurs d'un tel réseau ;

- les figures 3a et 3b illustrent schématiquement une architecture de base fonctionnant suivant l'invention ;

- la figure 4 illustre schématiquement un exemple de réalisation d'une architecture d'un système informatique complet du type réseau conforme à l'invention ;

- la figure 5 illustre schématiquement une topologie de bus utilisée dans le système selon la figure 4 ;

- La figure 6 illustre l'échange de données entre deux unités de traitement de données particulières du système informatique de la figure 4 ;

- la figure 7 est un chronogramme illustrant l'échange de données entre les deux unités de la figure 6 ;

- la figure 8 illustre une variante du système selon l'architecture de la figure 4.

La figure 1 illustre un exemple d'architecture d'un système informatique de type réseau conforme à l'art connu. Pour simplifier le dessin, on n'a représenté que deux unités connectées à un réseau local du type "LAN" ("Local Aera Network"). Ce réseau peut être de tout type et de toute configuration : bus (par exemple du type "ETHERNET") ou en anneau (par exemple du type "TOKEN RING"). Sur la figure 1, on a supposé qu'il s'agissait de deux serveurs, S_A et S_B, respectivement. Le serveur S_A pourrait être remplacé par une station de travail ou un simple micro-ordinateur. Ces serveurs, S_A et S_B, sont connectés physiquement au bus de type LAN via des interfaces classiques, I_A et I_B, qui dépendent de la nature du bus. Elles comprennent notamment des moyens de mémoire tampon et des moyens de codage et de décodage permettant, en particulier, de reconnaître une adresse de destination. On a supposé que chaque serveur était associé à des moyens de mémoire de masse, illustrés par les disques durs D_A ou D_B, connectés au serveur correspondant à l'aide de canaux d'entrée/sortie, I/O_A ou I/O_B,

respectivement. Dans la suite du texte, pour simplifier, on désignera ce réseau par le sigle LAN.

Il existe deux méthodes principales d'échange de données.

En local, l'unité de traitement de données, par exemple le serveur S_A, qui veut accéder à des adresses de mémoire de son propre disque, D_A, en lecture ou en écriture, se sert d'instructions élémentaires agissant directement sur les têtes de lecture/écriture du disque. Ces instructions élémentaires sont du type "Entrée/Sortie" connues sous la dénomination anglo-saxonne "I/O Channel". On peut parler de protocole de type "adressage de mémoire". L'accès est quasi instantané. Il dépend des caractéristiques de l'unité de disque (vitesse d'accès moyen, débit, etc.) et du mode utilisé, celui-ci étant par exemple du type connu et standardisé (aussi bien à l'ANSI qu'à l'ISO) sous la dénomination "SCSI" pour "Small Computer System Interface". Sur la figure 1, on a illustré la lecture d'un fichier F_A à partir du disque D_A, selon ce mode d'échange de données. Les données relatives à ce fichier F_A sont lues directement sur des pistes et secteurs du disque B_A, à des adresses déterminées enregistrées dans une table d'allocation.

Par contre, si l'on désire lire des données, par exemple un fichier F_B stocké sur une unité de stockage externe, en l'occurrence le disque D_B attaché au serveur S_B, il est nécessaire de passer par le réseau LAN, les serveurs, S_A et S_B, et les interfaces, I_A et I_B. Le serveur S_A ne peut plus commander directement les têtes du disque dur D_B. Les échanges s'effectuent selon un mode de communication par paquets de données, faisant usage de messages et mettant en oeuvre un protocole particulier propre au type de réseau local utilisé.

Le chronogramme de la figure 2 illustre schématiquement un tel échange de données par le réseau LAN. Le temps total nécessaire à un échange d'un paquet de données élémentaires est au moins égal à l'intervalle de temps T, qui se décompose en intervalles de temps élémentaires T_1 à T_5. L'intervalle de temps T_1, entre les instants arbitraires t=0 et $t=t_1$, correspond au temps de traitement dans le serveur S_B. L'intervalle de temps T_2, entre les instants $t=t_1$ et $t=t_2$, correspond aux pertes de temps dues au protocole de communication particulier utilisé et aux conditions de transmission instantanées régnant sur le réseau (charge, nombres d'unités connectées, etc.). L'intervalle de temps T_3, entre les instants $t=t_2$ à $t=t_3$, représente le temps de transmission. Il dépend

essentiellement de la distance entre deux stations connectées (éloignement des serveurs S_A et S_B dans l'exemple), du débit de la liaison et, à un degré moindre, de la nature physique de la liaison (vitesse de propagation des signaux). De nouveau, l'intervalle de temps T_4, entre les instants $t=t_3$ et $t=t_4$, représente la contribution du protocole de communication (à l'arrivée). Enfin, l'intervalle de temps T_5, entre les instants $t=t_4$ et $t=t_5$, représente le temps de traitement à l'arrivée.

On conçoit aisément que ce mode d'échange est beaucoup plus lent que le mode précédent. Outre les données utiles (par exemple celles relatives au fichier F_B), des données supplémentaires doivent être transmises, notamment des données propres au protocole et des données d'adresses (origine et cible). Pour tenir compte des erreurs, il est également nécessaire de transmettre des données redondantes (parité, codes de détection et/ou de correction d'erreur). En outre, il est nécessaire de tenir compte des contentions et des collisions possibles, sur le bus ou l'anneau, selon les cas. Pour pallier à ces problèmes, des méthodes particulières telles que "CSMA/CD", de l'anglo-saxon "Carrier Sense Multiple Access / Collision Detection" ou "Ecoute Porteuse, Accès Multiples / Détection Colision" ont été proposées ; elles font l'objet des recommandations de la norme IEEE 802.3. Ces phénomènes contribuent également à augmenter le temps moyen de transmission.

Selon une caractéristique principale de l'invention, on fait appel à un stockage distribué des données d'informations sur les différentes ressources de stockage d'un système informatique. Pour ce faire, on attribue à chaque serveur un espace mémoire virtuel de très grande capacité, incorporant son ou ses propres disque(s), s'il(s) existe(nt), et des disques externes et/ou d'autres ressources de stockage de données. Le serveur précité est alors à même d'adresser directement la totalité de l'espace mémoire virtuel qui lui est attaché, par des instructions du type "I/O channel" précédemment décrit. En d'autres termes, il commande des têtes de lecture/écriture virtuelles.

Les figures 3a et 3b illustrent schématiquement l'architecture de base conforme à l'invention.

Sur la figure 3a, un serveur S_X, qui peut par ailleurs être connecté au réseau local LAN (figure 1), est associé comme précédemment à une unité locale de stockage de données, par exemple un disque D_X. Selon un aspect important de

l'invention, le serveur communique avec le disque au travers d'une unité physique ou logique DSM_X, qui constitue les moyens d'administration distribués du système selon l'invention.

Dans le premier cas, quand l'unité DSM_X est du type physique, il s'agit de circuits de traitement de données à programme enregistré connectés physiquement au disque D_X, par un canal d'entrée/sortie I/O_X, et à des disques externes, par exemple aux disques D_Y et D_Z, par un système de bus B spécifique qui sera détaillé ci-après. Ce bus B permet de transmettre des instructions de type "I/O Channel" également aux disques D_Y et D_Z précités. Un programme spécifique, enregistré dans l'unité physique DSM_X, d'un type appelé ici "agent intelligent", permet ce fonctionnement particulier. A titre d'exemple non limitatif, l'unité DSM_X peut être constituée par une station de travail intermédiaire fonctionnant sous le système d'exploitation "UNIX" (marque déposée).

Dans le second cas, quand l'unité DSM_X est du type logique, le programme précité est directement enregistré dans la mémoire vive du serveur S_X, et c'est ce dernier qui est physiquement connecté aux différents disques, D_X à D_Z, le serveur jouant son rôle propre et celui des moyens d'administration distribués (DSM_X). Il s'agit donc d'une unité purement logique, intégrée à la station de travail S_X.

Dans une autre variante de réalisation, l'unité logique peut être intégrée directement dans les circuits électroniques associés aux ressources de stockage, par exemple dans le contrôleur, s'il s'agit d'un disque.

Comme illustré plus particulièrement par la figure 3b, le serveur S_X "voit" l'ensemble des disques D_X à D_Z comme un seul disque virtuel, D'_{XYZ}, de grande ou très grande capacité. En l'occurrence, sa capacité est au moins égale à la capacité cumulée des disques D_X à D_Z. Naturellement, la capacité totale du disque virtuel doit être dimensionnée de façon à être adaptée à l'espace disque maximum adressable par l'unité arithmétique et logique du serveur S_X. Il s'ensuit que le serveur S_X adresse directement des pistes ou secteurs S_m du disque virtuel précité D'xyz, comme si ce disque était un disque local. Il envoie des instructions de déplacement de têtes virtuelles et des ordres élémentaires d'écriture et lecture comme pour un disque local, selon le mode de fonctionnement "I/O Channel" précité.

Comme il sera détaillé ultérieurement, dans un mode préféré de l'invention, plusieurs type "d'agents intelligents" sont mis en oeuvre. Pour les besoins d'adressage précités, on peut appeler ces agents particuliers "agents de stockage". Il s'agit de pièces de programme chargées dans une mémoire vive de l'unité physique DSM_X, ou alternativement du serveur S_X lorsque DSM_X est une unité logique intégrée dans le serveur S_X.

Le rôle de ces programmes élémentaires est double :

a/ aiguillage vers une unité de stockage particulière, soit locale, par exemple D_X, soit distante, par exemple D_Y et/ou D_Z, selon la piste virtuelle adressée et :

b/ adaptation éventuelle du mode d'adressage et traduction de protocole, en fonction de l'unité physique réelle, D_Y et D_Z.

En effet, selon un aspect avantageux de l'invention, le serveur S_X peut adresser des unités de stockage hétérogènes (technologies et/ou constructeurs différents, mode de fonctionnement différents, etc.). A titre d'exemple non limitatif, le disque Dx pourrait être un disque mémorisant des mots codés "ASCII" ("American Standard Code for Information Interchange") de longueur 32 bits, le serveur S_X fonctionnant sous le système d'exploitation "UNIX" et les disques D_Y et D_Z, étant des disques attachés à un ordinateur de grande puissance dit "Main Frame", mémorisant des mots codés "EBCDIC" ("Extended Binary Coded Decimal Interchange Code") de longueur 36 bits.

L'unité physique DSM_X ou son équivalent logique, si elle est intégrée dans le serveur S_X, rend les opérations d'adressage du disque D'_{XYZ} entièrement transparentes, que ce soit le disque local D_X ou les disques externes D_Y ou D_Z qui stockent physiquement l'information. De façon plus précise, c'est l'agent intelligent spécialisé enregistré, ou agent intelligent de stockage, dans ce cas précis, qui accomplit cette tâche. Le serveur S_X adresse le disque virtuel D'_{XYZ} par l'intermédiaire d'un canal d'entrée-sortie virtuel I/O'_X, comme si c'était son propre disque local D_X et selon le ou les seul(s) protocole(s) qu'il "connaît", par exemple celui associé au disque local D_X.

La figure 4 illustre schématiquement un exemple de système d'informatique 1 à architecture réseau, incluant des dispositions propres à l'invention (représentées en grisé sur la figure).

On suppose que le réseau local LAN est du type "INTRANET". Ce type de réseau intègre, en local, les technologies propres au réseau "INTERNET". Ce réseau comprend des serveurs faisant appel au protocole de communication de base connu sous le sigle "TCP/IP". Il inclut aussi d'autres protocoles de communication tels que "HTTP", "NFS", etc., également utilisés sur le réseau "INTERNET". Tous ces protocoles sont normalisés.

On a d'ailleurs supposé que le réseau local LAN communique avec le réseau "INTERNET" I_T, par exemple par l'intermédiaire de l'un des serveurs énumérés ci-dessous.

A ce réseau local LAN sont connectés, d'une part des stations de travail, ST_1 à ST_3, d'autre part des serveurs de stockage S_1 à S_3, dans l'exemple décrit. Ces serveurs de stockage, S_1 à S_3, seront appelés dans ce qui suit "serveurs", par esprit de simplification.

Selon l'une des caractéristiques principales de l'invention, les ressources de stockage, attachées à chaque serveur, S_1 à S_3, le sont via une unité d'administration distribuée, physique ou logique, DSM_1 à DSM_3. Dans l'exemple décrit, il s'agit de deux unités de disques classiques, D_1 à D_3.

Le système informatique 1 peut comprendre d'autres unités de stockage, telles qu'une librairie de bandes magnétiques ou de cartouches magnétiques, TL_4, qu'un ensemble de disques, FD_6, appelé "ferme de disques" ("Disk Farms", selon l'appellation anglo-saxonne généralement utilisée), ou encore que des ressources de stockage éloignées, représentées sous la référence générale STO_e. Ces ressources de stockages sont gérées également par des unités d'administration qui leur sont affectées, DSM_4, DSM_6 et DSM_5, respectivement. La ressource de stockage éloignée, STO_e, communique avantageusement avec l'unité DSM_5, via une liaison, I_{ATM}, en mode "ATM" (de l'anglo-saxon "Asynchronous Transfer Mode", ou "Mode de Transfert Asynchrone").

Selon un autre aspect de l'invention, les différentes unités d'administration distribuées, DSM_1 à DSM_6, sont connectées entre elles via un bus B à très grande vitesse. Elles sont également reliées, toujours via ce bus B, à une unité d'administration centralisée NSM.

Cette dernière unité, NSM, qui peut être constituée à base d'une station de travail, par exemple fonctionnant sous le système d'exploitation "UNIX", comprend des moyens de mémoire, par exemple un disque, emmagasinant une base de données CDB. Ces données comprennent la description du système, et notamment de toutes les ressources de stockage et de leurs caractéristiques (capacité, mode de fonctionnement, protocoles, etc.). Elle comprend également des données décrivant l'affectation de ces ressources de stockage aux différents serveurs, S_1 à S_3, du système informatique 1. Il doit être clair que chaque ressource de stockage, et plus particulièrement la librairie de bandes magnétiques, TL_4, la "ferme de disques", FD_6, et/ou les ressources de stockage éloignées, STO_e, peut être partagée entre plusieurs serveurs, S_1 à S_3. Ce partage peut s'effectuer sur une base dynamique, en ce sens qu'il n'est pas figé une fois pour toutes et/ou qu'il dépend des applications en cours de traitement. La librairie de bandes ou de cartouches magnétiques peut notamment servir à des opérations de sauvegarde périodiques de tout ou partie des données enregistrées sur les disques du système 1, disques attachés directement ou indirectement à un ou plusieurs serveurs, S_1 à S_3.

A partir des données précitées de la base de données enregistrée CDB, l'unité d'administration centralisée NSM élabore de façon connue les programmes spécialisés précités, ou "agents intelligents". Pour ce faire, selon un aspect avantageux de l'invention, on met à profit une technologie de programmation du type "JAVA-Applets", par ailleurs utilisée en conjonction avec le réseau "INTERNET". Il s'agit d'un langage orienté objet et d'un environnement de type "run-time", c'est-à-dire dont les programmes peuvent s'auto-exécuter à réception sur l'unité cible. En effet, l'exécution de ces programmes ne dépend pas de l'environnement de réception (système d'exploitation sous "UNIX", "WindowsNT", "Windows 95", [marques déposées], etc.). On parle de "Machine Virtuelle JAVA". Les applications "JAVA" peuvent donc s'exécuter sur toutes les unités où un logiciel "Machine Virtuelle JAVA" est implanté. L'exécution des programmes est donc indépendante de la plate-forme utilisée. Enfin, les communications s'effectuent en mode "Client - Serveur".

L'invention tire partie de ces caractéristiques avantageuses. Les "agents intelligents" sont programmés en langage "JAVA" et transmis dynamiquement et sélectivement, via le bus à haut débit B, aux différentes unités d'administration distribuées, DSM_1 à DSM_6, qu'elles soient physiques ou logiques (c'est-à-dire, dans ce dernier cas, confondues avec les serveurs). A réception, les programmes

s'auto-exécutent dans les unités DSM_1 à DSM_6. Cette exécution se traduit pratiquement par le téléchargement et la mémorisation d'instructions en mémoire vive de ces unités (ou des serveurs lorsque les unités DSM_1 à DSM_6 sont logiques).

Ces instructions attribuent notamment, à un instant donné, un espace mémoire virtuel (par exemple, figure 3b : D'_{xyz}) au serveur (par exemple S_X) associé à une unité d'administration distribuée donnée (par exemple DSM_X).

Lorsqu'un serveur, par exemple S_1, émet une requête pour la lecture et/ou l'écriture de données dans l'espace mémoire virtuel qui lui est attribué, l'unité d'administration distribuée adresse, sous la commande des instructions téléchargées, soit le disque local D_1, soit un disque externe, par exemple D_3. Dans tous les cas, l'adressage est effectué en mode "I/O Channel" et non sous la forme d'un protocole de communication, à l'aide de messages. Pour le disque local, D_1, le protocole d'adressage est celui utilisé par le serveur S_1. Pour le disque externe, D_3, l'adressage s'effectue par ce même protocole, via le bus B et l'unité d'administration distribuée, DSM_3, associée au disque D_3. Cette dernière unité, DSM_3, doit éventuellement traduire le protocole d'adressage de la manière décrite, de façon à s'affranchir des hétérogénéités des matériels utilisés. L'opération d'adressage finale, c'est-à-dire la commande des têtes de lecture-écriture du disque physique D_3, est effectuée sous la commande des instructions téléchargées et mémorisées dans l'unité d'administration distribuée DSM_3. Ces instructions sont issues, comme précédemment, de l'auto-exécution dans DSM_3 d'un "agent intelligent de stockage" transmis par l'unité d'administration centralisée NSM.

On doit comprendre également qu'une requête d'écriture et/ou de lecture, par un serveur, dans son espace disque virtuel, peut se traduire physiquement, en bout de chaîne, par la lecture ou l'enregistrement de données sur un autre support, une bande ou une cartouche magnétique par exemple. C'est le cas d'une requête lancée par le serveur S_2 par exemple, qui aboutirait à l'unité DSM_4 et à la librairie de bandes magnétiques TL_4. Le changement fondamental du type d'adressage qui en résulte reste transparent pour le serveur S_3. Ce sont les instructions téléchargées et mémorisées dans DSM_4 qui effectuent les traductions de protocoles et adaptations nécessaires. Il est clair que, dans ce cas, DSM_4 n'étant pas connectée à un serveur, il ne peut s'agir que d'une unité physique et non logique.

Pour que le système puisse remplir les exigences du procédé de l'invention, il est nécessaire que le bus B accepte un protocole du type "I/O Channel" précité. En d'autres termes, les échanges sur le bus ne s'effectuent pas selon un protocole de type communication.

Pour ce faire, selon un aspect de l'invention on choisit un bus au standard dit "Fibre Channel" qui a fait l'objet de normes (ANSI X3.230, de 1994). Ce type de bus est encore appelé "Fibre Backbone", selon la terminologie anglo-saxonne. Il s'agit d'un bus pouvant véhiculer des données à très haut débit. Le but principal assigné à un tel bus est de transmettre les données d'un point à un autre sous un temps d'attente très faible. Seule une correction d'erreur simple est effectuée, ce par le matériel et non par le logiciel. Les données sont transmises dans les deux directions, simultanément. Lorsqu'une transmission échoue pour cause de congestion, elle est ré-initiée immédiatement sans intervention logicielle. Enfin, ce type de bus est compatible avec les protocoles de haut niveau tels "SCSI" précédemment signalé. Il permet donc de véhiculer des requêtes de type "I/O Channel".

Il existe trois topologies de bus possibles : du type "Point-à-Point", du type dit "Switched Fabric" et du type "Anneau avec arbitrage" ("Arbitred Loop"), similaire aux anneaux "à jetons" ("TOKEN RING").

Dans le cadre de l'invention, on choisit de préférence la seconde topologie, qui offre la plus grande capacité de connexion. La figure 5 illustre schématiquement une telle topologie. Selon cette topologie, chaque dispositif, c'est-à-dire selon l'invention chacune des unités d'administration distribuées, DSM_1, ..., DSM_n à DSM_X, et centralisée, NSM, est connecté à un commutateur et reçoit une voie de données non bloquante pour n'importe quelle autre connexion sur le commutateur. Cette disposition est équivalente à une connexion dédiée avec n'importe quelle autre unité. Quand le nombre d'unités augmente et occupe une multitude de commutateurs, ces commutateurs sont à leur tour connectés entre eux. Il est recommandé d'établir des voies de connexion multiples entre commutateurs pour créer une redondance de circuits et augmenter la bande passante totale.

Bien que le standard prévoie la possibilité d'utiliser différents supports physiques pour réaliser les liaisons (paire torsadée, câble coaxial miniature ou vidéo, fibre optique multimode ou monomode), on choisit, dans le cadre de

l'invention, la fibre monomode. Ce type de support permet, à la fois un très haut débit (100 MO/s) et des liaisons à grande distance (jusqu'à 10 km).

Les choix ci-dessus permettent donc d'établir des liaisons allant typiquement de 10 m à 10 km, pour des débits allant jusqu'à 100 MO/s, et pouvant accepter des centaines de connexions en parallèle aux débits précités. Enfin, le bus B est compatible avec de nombreux protocoles, qu'ils soient de type communication comme le protocole "TCP/IP" ou de haut niveau du type "I/O Channel" ("SCSI" par exemple). Le bus B est équivalent à une multitude de ports de tye "I/O Channel".

Ceci permet de réduire très significativement le temps nécessaire aux échanges de données entre deux serveurs, dans un rapport typique de un à dix. La figure 6 illustre schématiquement des échanges de données entre deux serveurs : un ordinateur MF du type "Main Frame" et une station de travail STX fonctionnant sous le système d'exploitation "UNIX". Selon la caractéristique principale de l'invention, les disques durs locaux, D_{MF} et D_{STX}, respectivement, sont attachés aux serveurs via des unités d'administration distribuées, DSM_{MF} et DSM_{STX}, respectivement. On suppose que les deux disques sont partagés entièrement entre les deux serveurs. Chaque serveur voit donc l'espace disque total comme un seul disque virtuel D_V de capacité égale à la somme des deux disques.

On suppose que l'un des serveurs émet une requête, par exemple que le serveur MF émet une requête d'écriture ou de lecture de données, sur le disque virtuel D_V. Cette requête est transmise à l'unité DSM_{MF} pour exécution. Si l'espace disque physique concerné n'est pas dans l'espace disque local D_{MF}, la requête va être transmise à l'unité DSM_{STX}, via le bus B. Celle-ci va, dans l'exemple précis, effectuer deux opérations : traduction de protocoles (puisque, comme indiqué précédemment, les modes d'enregistrement de données sont différents sur les deux disques : "EBCDIC" et "ASCII" respectivement, largeurs de mots différents) et adressage physique du disque D_{SXT} à partir de l'adresse virtuelle transmise.

Le chronogramme de la figure 7 détaille le temps nécessaire à l'échange précité. L'intervalle de temps total T' ne comporte plus que trois phases : deux phases extrêmes de traitement : T'_1 et T'_3, entre les instants respectifs t=0 et $t=t'_1$, d'une part, et $t=t'_2$ et $t=t'_3$, d'autre part, et une phase de transmission via le bus B : T'_2, entre les instants $t=t'_1$ et $t=t'_2$.

Les intervalles de temps T'_1 et T'_3 doivent être comparés aux intervalles de temps T_1 et T_5 de la figure 2 (art connu). A priori, pour des matériels et des traitements/applications donnés identiques, ces intervalles de temps sont égaux. Ils ne dépendent que des conditions locales de traitement de l'information et en aucune façon des protocoles de transmission, ni du temps de transmission.

Par contre, il n'y a plus, contrairement à l'art connu, de pertes de temps en relation avec le protocole de communication : intervalles de temps T_2 et T_4 (figure 2). Ceci est dû aux propriétés du bus B, qui est compatible avec les protocoles de haut niveau ("SCSI" par exemple). Enfin, puisque l'on utilise un bus à très haut débit et la topologie dite "Switched Fabric", le temps de transmission est réduit à son strict minimum. L'intervalle de temps T'_2 est donc très inférieur à l'intervalle de temps homologue T_3 (figure 2).

L'intervalle de temps total T' nécessaire à un échange se réduit donc essentiellement aux intervalles de temps de traitement en local et devient pratiquement indépendant des transmissions entre unités. En d'autres termes, les échanges entre une ressource de stockage locale et un serveur, ou entre une ressource de stockage externe et ce même serveur, s'effectuent à une vitesse, sinon identique, du moins très comparable. Les performances du système ne sont donc pas dégradées, quelle que soit la localisation de la ressource de stockage de données. En d'autres termes encore, n'importe quel serveur, S_1 à S_3 (figure 4), du système informatique 1, peut, a priori, accéder à n'importe quelle ressource de stockage, D_1 à D_3, TL_4, FD_6 ou STO_e, sans risque de dégradation des performances.

Il s'ensuit aussi que les données d'informations peuvent être réparties de façon optimisée dans les différentes ressources de stockage du système informatique, sur un même site ou sur un site éloigné (figure 4 : STO_e), dans la mesure où le temps de transmission (figure 6 : T'_2) reste dans des limites acceptables.

La répartition optimisée précitée peut s'effectuer en tenant compte de nombreux critères : performances (vitesses) et/ou capacités des différentes ressources de stockage, types de données, défaillances momentanées d'une ou plusieurs ressources de stockage, en tout ou partie, dépassement de capacité, coût du stockage, etc.

La gestion du stockage distribué est réalisée, comme il a été indiqué, par l'unité d'administration centralisée, en tenant compte de la base de donnée CDB. Celle-ci élabore et télécharge dynamiquement, dans les différentes unités d'administration distribuées, DSM_1 à DSM_6, des agents intelligents de stockage. Les agents de ce type véhiculent des instructions tenant compte, notamment : des conditions locales (type de matériel connecté, capacité de mémoire locale, protocole utilisé, etc.), de l'espace disque virtuel total attribué à un serveur donné, et de l'adresse de l'unité d'administration centralisée ou des unités d'administration distribuée(s) qui gère(nt) la ou les ressource(s) de stockage physique(s). Ces caractéristiques varient en fonction du temps. Aussi, les opérations effectuées sous la commandes des agents sont dynamiques.

De façon pratique, la gestion s'effectue de façon similaire à la gestion d'un réseau local de transmission de données classique. L'unité NSM d'administration centralisée, de ce qui peut être appelé un "réseau" d'unités d'administration distribuées, DSM_1 à DSM_6, peut être concrétisée par une unité de traitement de données telle qu'un serveur sous "UNIX", voire être confondue avec l'unité d'administration du réseau local classique LAN, coexistant avec le bus B.

Outre les agents intelligents de stockage, d'autres types d'agents peuvent être mis en oeuvre avec profit dans le cadre de l'invention.

Un premier type d'agents supplémentaires est constitué par des agents de gestion d'opérations d'entrée-sortie spéciales. Ces agents sont téléchargés dans certains noeuds de stockage, c'est-à-dire pratiquement dans les unités d'administration distribuées associées à ces noeuds, pour obtenir des résultats spécifiques sur les opérations d'entrées-sorties précitées. Il peut s'agir par exemple d'optimiser les temps d'accès à des bases de données.

Un deuxième type d'agents supplémentaires est constitué par des agents organisant des sauvegardes automatiques en fonction d'un certain nombre de paramètres : date, heure, type de données, etc. Ces agents organisent le déplacement de données, par duplication, vers des moyens de stockage de données prédéterminés.

Un troisième type d'agents supplémentaires est constitué par des agents gérant l'archivage et le stockage hiérarchique de données entre les ressources de stockage présentes dans le système. Ils permettent notamment de

déplacer les données d'une ressource à l'autre, en tenant compte de contraintes telles que temps d'accès, fréquence d'accès, localisation et coût. Plus particulièrement, un agent spécifique est affecté à l'unité d'administration distribuée DSM_4 attachée aux librairies de bandes ou cartouches magnétiques TL_4, que l'on peut appeler "agent serveur de média". Il apporte une vision virtuelle des objets de stockage physiques et peut accommoder des requêtes d'entrée-sortie formulées par des serveurs externes. Ces derniers ont accès à ces objets de stockage en émettant des requêtes de lecture-écriture classiques, comme s'ils avaient affaire à leurs ressources de stockage locales, par exemple un disque. Grâce à cette disposition, les serveurs d'applications pourront tirer partie de n'importe quelle technologie de stockage de masse, que cette technologie soit actuellement opérationnelle ou en cours d'élaboration, ce sans modification des applications.

Un quatrième type d'agents est constitué par des agents de gestion des moyens d'enregistrement en redondance, en fonction du profil des données, des applications spécifiques et du coût induit, par exemple. Pour des raisons de sécurité, notamment intégrité et disponibilité des données, celles-ci sont enregistrées avec une redondance plus ou moins accentuée, voire dupliquées. On adjoint aussi, aux données utiles d'information, des données de redondance : clés de parité, codes de détection d'erreurs ("EDC" pour "Error-Detection Code") ou de correction d'erreurs ("ECC" pour "Error-Correcting Code"). Ces dispositions conduisent à diverses techniques d'enregistrement : disques dits "miroirs" ou "RAID5", c'est-à-dire avec calcul de clé de parité. Les agents de ce quatrième type gèrent de façon optimisée l'utilisation d'une technique ou d'une autre.

Un cinquième type d'agents est constitué par des agents gérant les connexions avec les sites éloignés de stockage de données STO_e. Comme il a été indiqué, les liaisons s'effectuant en mode "ATM", il est donc nécessaire d'effectuer une traduction de protocole : "Fibre Channel" à "ATM". L'agent de ce type est téléchargé dans l'unité d'administration distribuée DSM_5, en sus d'autres agents (agents de stockage, etc.).

Un agent d'un sixième type peut être téléchargé dans une des unités d'administration distribuées pour gérer les échanges avec le réseau "INTERNET" et aiguiller les données téléchargées par ce moyen vers une ou plusieurs ressources de stockage du système informatique 1.

Un septième type d'agents est constitué par des agents surveillant et gérant le bon fonctionnement de l'ensemble des ressources de stockage de données.

D'autres types d'agents peuvent être élaborés, en tant que de besoin, par l'unité d'administration centralisée NSM, et télédéchargés sélectivement dans les unités d'administration distribuées, via le bus B. De tels agents sont notamment nécessaires à chaque fois qu'il s'agit de traduire un premier protocole en un second.

Il a été également signalé que, dans les problèmes liés à la sécurité de traitement de données, la confidentialité devait être prise en compte. Cet aspect peut être traité également en faisant appel à des agents spécialisés. Ces agents peuvent vérifier, par exemple, si les requêtes sont légitimes et/ou réservées, en fonction du contenu des données, du profil des utilisateurs ou des ressources accédées.

Dans l'architecture présentée sur la figure 4 coexistent deux bus : un bus local en anneau (dans l'exemple décrit) LAN et un bus spécifique à l'invention, le bus B à haut débit. Selon cette architecture, les communications entre les serveurs de stockage, S_1 à S_3, et les autres unités, par exemple les stations de travail ST_1 à ST_3, s'effectuent de façon classique sous forme de transferts de fichiers, en faisant appel à des protocoles de communication compatibles avec le type de réseau local LAN.

Cependant, comme le bus B, conforme au standard "Fibre Channel", est susceptible de véhiculer plusieurs protocoles, y compris des protocoles de type communication, il est tout-à-fait possible de supprimer le réseau LAN (figure 4) et de faire transiter toutes les transmissions par le seul bus B.

La figure 8 illustre schématiquement une architecture de système informatique 1' ne comportant qu'un seul type de bus, le bus B. Toutes les unités, qu'elles soient ou non spécifiques à l'invention, sont connectées à ce bus B. A titre indicatif, on a représenté sur cette figure 8 un serveur S_1 attaché à un disque local D_1 via une unité d'administration distribuée DSM_1, un deuxième serveur S_7 gérant les communications avec le réseau "INTERNET" IT et connecté au bus B via une unité d'administration distribuée DSM_7, une unité d'administration distribuée DSM_4 attachée à la librairie de bandes ou cartouches magnétiques TL_4, l'unité

d'administration centralisée NSM et sa base de données représentée par un disque CDB, ainsi que trois stations de travail, ST_1 à ST_3, désormais connectées directement au bus B. Les stations de travail, ST_1 à ST_3, communiquent entre elles de façon classique, c'est-à-dire en faisant usage d'un protocole de communication. En effet, comme il a été indiqué, les deux types de protocole peuvent coexister sur le bus B. Dans cet esprit, les serveurs, par exemple le serveur S_1 peut être connecté aussi directement au bus B et communiquer avec d'autres unités (par exemple ST_1 à ST_3) sous un protocole de communication, par exemple "TC/IP".

A la lecture de ce qui précède, on constate aisément que l'invention atteint bien les buts qu'elle s'est fixés.

Elle permet notamment d'offrir à chaque serveur une vue virtuelle d'un espace de stockage global. A priori, elle permet à n'importe quel serveur d'accéder à n'importe quelle ressource de stockage du système, ce qu'elle qu'en soit la nature. Elle permet, en outre, d'accéder à des ressources éloignées, sans dégradation substantielle des performances.

Il doit être clair cependant que l'invention n'est pas limitée aux seuls exemples de réalisation précisément décrits, notamment en relation avec les figures 3a à 8. Elle s'accommode de nombreux types de matériels. Ceci est vrai en premier lieu pour les unités centrales de traitement de données, qui peuvent comprendre indifféremment des micro-ordinateurs, des stations de travail, des mini-ordinateurs, ou des ordinateurs de grande puissance du type dit "main frames". Les systèmes d'exploitation peuvent également être divers et, par exemple, comprendre des systèmes d'exploitation dits universels ou propriétaires. Outre le bus B, il peut être fait usage de réseaux locaux divers : bus (du type "ETHERNET", par exemple) ou en anneau (du type "TOKEN RING", par exemple). Enfin, et surtout, le type de périphériques de stockage n'est pas limité : des unités de disques magnétiques, de disques optiques, de bandes et/ou de cartouches magnétiques, sur site local ou éloigné, peuvent être mises en oeuvre.

REVENDICATIONS

1. Système informatique (1) comprenant une pluralité de moyens de stockage de données distribués (D_1-D_3, FD_6, TL_4, STO_e) et au moins un serveur de stockage desdites données (S_1-S_3), caractérisé en ce qu'il comprend des moyens d'attribution à chacun desdits serveurs de stockage de données (S_1-S_3) d'un espace mémoire virtuel (D'_{xyz}) dont l'étendue est au moins égale à la capacité cumulée de tout ou partie desdits moyens de stockage de données distribués (D_1-D_3, FD_6, TL_4, STO_e).

2. Système selon la revendication 1, caractérisé en ce que lesdits moyens d'attribution sont constitués par une série de premiers moyens de traitement de données (DSM_1-DSM_6), à programmes enregistrés, dits moyens d'administration distribués, gérant chacun une partie desdits moyens de stockage de données (D_1-D_3, FD_6, TL_4, STO_e).

3. Système selon la revendication 1 ou 2, caractérisé en ce qu'il comprend en outre des seconds moyens de traitement de données (NSM), dits moyens d'administration centralisés, associés à une base (CDB) de données descriptives d'au moins la configuration desdits moyens de stockage de données (D_1-D_3, FD_6, TL_4, STO_e), en ce que lesdits moyens d'administration centralisés (NSM) comprennent des moyens d'élaboration et de téléchargement des dits programmes à partir desdites données dans lesdits moyens d'administration distribués (DSM_1-DSM_6), de manière à ce que ceux-ci, lorsqu'ils sont associés à un serveur de stockage (S_1-S_3), attribuent à celui-ci, sous la commande dudit programme téléchargé, ledit espace mémoire virtuel (D'_{xyz}).

4. Système selon l'une des revendications 1 à 3, caractérisé en ce qu'au moins une partie desdits moyens de stockage de données (D_1-D_3, FD_6, TL_4, STO_e) sont attachés à un serveur de stockage de données (S_1-S_3), via l'un desdits moyens d'administration distribués (DSM_1-DSM_6), pour former des moyens de stockages de données locaux dont l'espace mémoire est adressable par un protocole comprenant des instructions d'entrée-sortie de type écriture-lecture émises par ledit serveur de stockage de données (S_1-S_3), en ce que

lesdits moyens de stockage de données locaux forment une première partition dudit espace mémoire virtuel (D'_{xyz}), et en ce que des moyens de stockage externes audit serveur de stockage de données forment au moins une deuxième partition de cet espace mémoire virtuel (D'_{xyz}).

5. Système selon l'une des revendications 1 à 4, caractérisé en ce qu'il comprend un bus à haut débit (B) du type comprenant une multitude de ports "entrée-sortie" autorisant la transmission dudit protocole comprenant des instructions d'entrée-sortie de type écriture-lecture, en ce que lesdits moyens d'administration distribués (DSM_1-DSM_6) et lesdits moyens d'administration centralisés (NSM) sont connectés entre eux par ce bus (B), et en ce que lesdites instructions d'entrée-sortie émises par ledit serveur de stockage (S_1-S_3) de données sont aiguillées par lesdits moyens d'administration distribués (DSM_1-DSM_6), sous la commande desdits programmes téléchargés, vers lesdits moyens de stockage locaux formant ladite première partition de l'espace mémoire virtuel (D'_{xyz}), ou, via ledit bus (B), vers des moyens d'administration distribués (DSM_1-DSM_6) attachés auxdits moyens de stockages externes formant une desdites partition supplémentaires de l'espace mémoire virtuelle (D'_{xyz}), selon que lesdites instructions concernent l'une ou l'autre de ces partitions, de manière à rendre directement adressable l'ensemble dudit espace mémoire virtuel (D'_{xyz}) attribué audit serveur de stockage de données (S_1-S_3).

6. Système selon la revendication 5, caractérisé en ce que le support de transmission composant ledit bus (B) est une fibre optique monomode et en ce que lesdits ports comprennent une multitude de commutateurs de manière à créer des voies de transmission directes à partir de chacun des moyens d'administration connectés vers tous les autres.

7. Système selon l'une des revendications 1 à 6, caractérisé en ce que lesdits moyens de stockage comprennent des unités de disques magnétiques (D_1-D_3, FD_6), des unités de bandes ou cartouches magnétiques (TL_4), ou des disques optiques.

8. Système selon l'une des revendications 1 à 7, caractérisé en ce que l'un au moins (STO_e) desdits moyens de stockage de données est localisé sur un site

éloigné, en ce que les transmissions entre ce site et ledit système informatique (1) s'effectuent par des liaisons (IATM) à grande vitesse, en mode asynchrone, et en ce que lesdites liaisons (IATM) sont connectées à un desdits moyens d'administration distribués (DSM_5), ces moyens étant eux-mêmes connectés audit bus (B).

9. Système selon l'une des revendications 1 à 8, caractérisé en ce qu'il est configuré de telle sorte que lesdits programmes élaborés par lesdits moyens d'administration centralisés (NSM) sont transmis par ledit bus (B) pour être téléchargés sélectivement dans lesdits moyens d'administration distribués (DSM_1-DSM_6), en ce que les programmes sont écrits dans un langage déterminé les rendant auto-exécutables lors dudit téléchargement, et en ce que lesdits moyens d'administration centralisés (NSM) gèrent dynamiquement ledit téléchargement en fonction de paramètres évoluant dans le temps et de traitements effectués par ledit système informatique (1).

10. Système selon la revendication 9, caractérisé en ce que lesdits programmes téléchargés comprennent des programmes de gestion de l'espace mémoire virtuel (D'_{xyz}) attribué à chaque serveur de stockage (S_1-S_3).

11. Système selon la revendication 9, caractérisé en ce que, lesdits moyens de stockage de données (D_1-D_3, FD_6, TL_4, STO_e) étant de nature hétérogène et fonctionnant selon des modes et/ou protocoles différents, lesdits programmes téléchargés comprennent des programmes de traduction de mode et/ou protocole, de manière à ce que chaque serveur de stockage (S_1-S_3) accède audit espace mémoire virtuel (D'_{xyz}) qui lui est attribué en mettant en oeuvre ses propres protocoles.

12. Système selon la revendication 9, caractérisé en ce que lesdits programmes téléchargés comprennent des programmes d'archivage automatique de données, selon une hiérarchie déterminée, dans les moyens de stockage de données distribués (D_1-D_3, FD_6, TL_4, STO_e).

13. Système selon la revendication 9, caractérisé en ce que lesdits programmes téléchargés comprennent des programmes de sauvegarde de données par stockage de celles-ci selon un schéma redondant déterminé.

14. Système selon la revendication 9, caractérisé en ce que, l'un au moins (STO_e) desdits moyens de stockage de données étant localisé sur un site éloigné, en ce que les transmissions entre ce site et ledit système informatique s'effectuent par des liaisons à grande vitesse (I_{ATM}), en mode asynchrone, lesdits programmes téléchargés comprennent des programmes de traduction de protocole de transmission de données, et en ce que ces programmes sont téléchargés dans des moyens d'administration distribués (DSM_5) formant interface entre ledit bus (B) et lesdites liaisons (I_{ATM}) en mode asynchrone.

15. Système selon l'une des revendications 1 à 13, caractérisé en ce que au moins une partie desdits moyens d'administration distribués (DSM_1-DSM_6) font partie intégrante du serveur de stockage de données (S_1-S_3) ou desdits moyens de stockage de données distribués (D_1-D_3, FD_6, TL_4, STO_e), avec lesquels ils sont associés pour former une unité logique de ceux-ci.

16. Système selon l'une des revendications 1 à 15, caractérisé en ce qu'il comprend des unités de traitement de données supplémentaires (ST_1-ST_3), un réseau local (LAN) et des moyens de connexion dudit système informatique (1) à un réseau de transmission de données externe (I_T), et en ce que lesdites unités supplémentaires (ST_1-ST_3), lesdits moyens de connexion et lesdits serveurs de stockage de données (S_1-S_3) sont connectés audit réseau local (LAN).

17. Système selon la revendication 16, caractérisé en ce que ledit bus (B) et ledit réseau local (LAN) sont confondus en un réseau de transmision de données unique (B), lesdites unités supplémentaires (ST_1-ST_3) étant connectées audit bus (B) et communiquant entre elles selon un protocole de communication via ce bus (B).

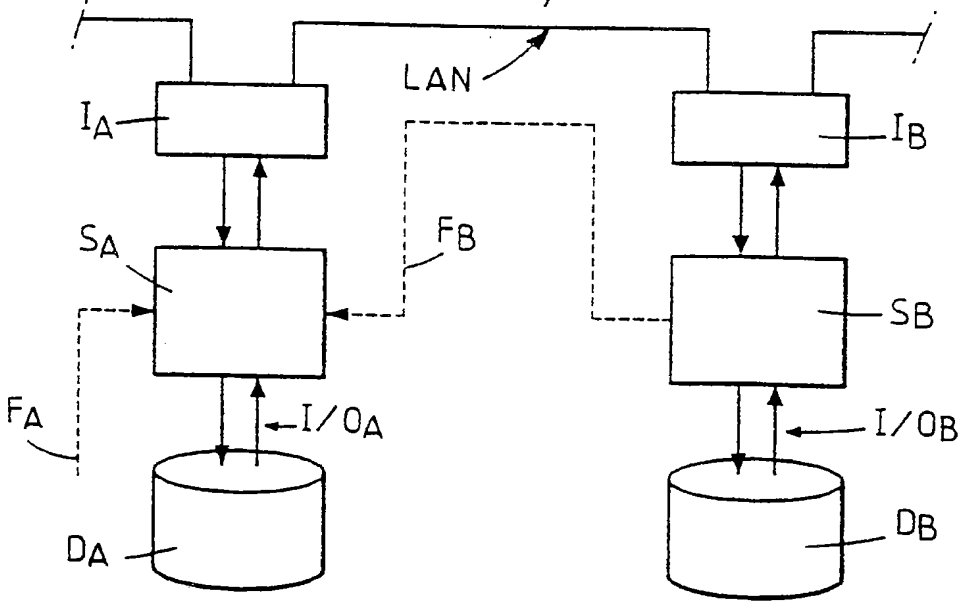


FIG. 1

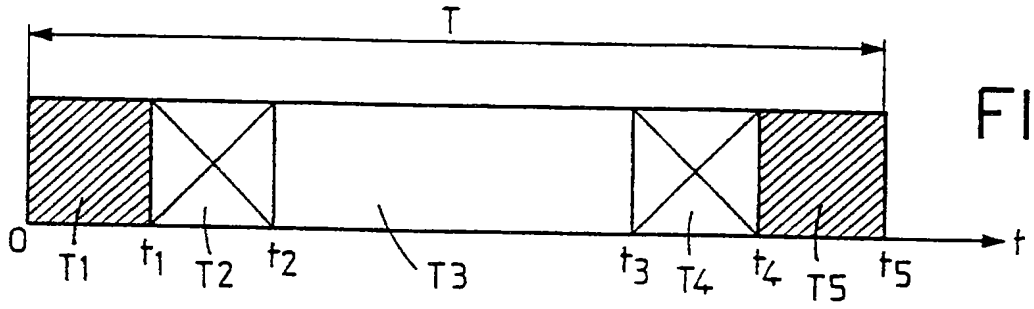


FIG. 2

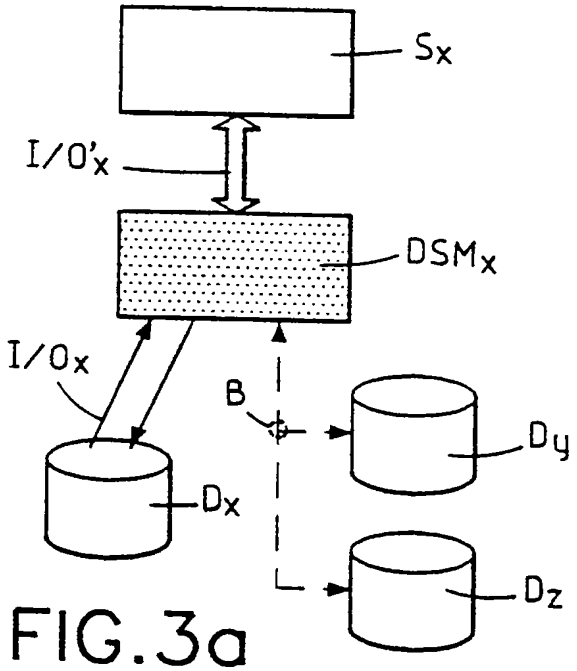


FIG. 3a

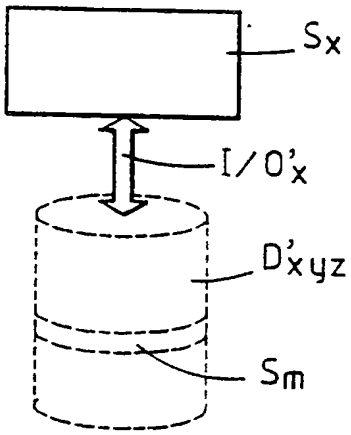


FIG. 3b

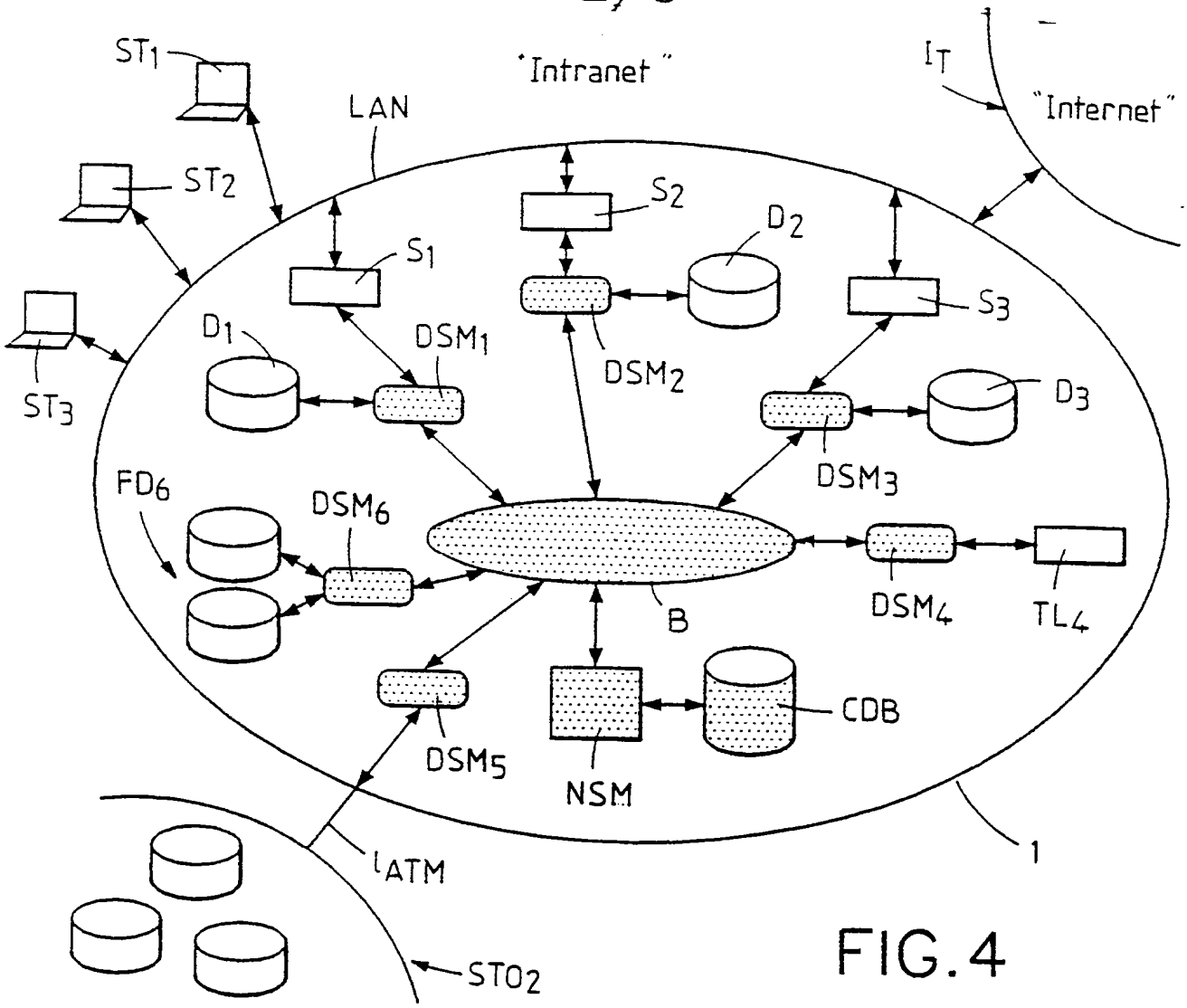


FIG.4

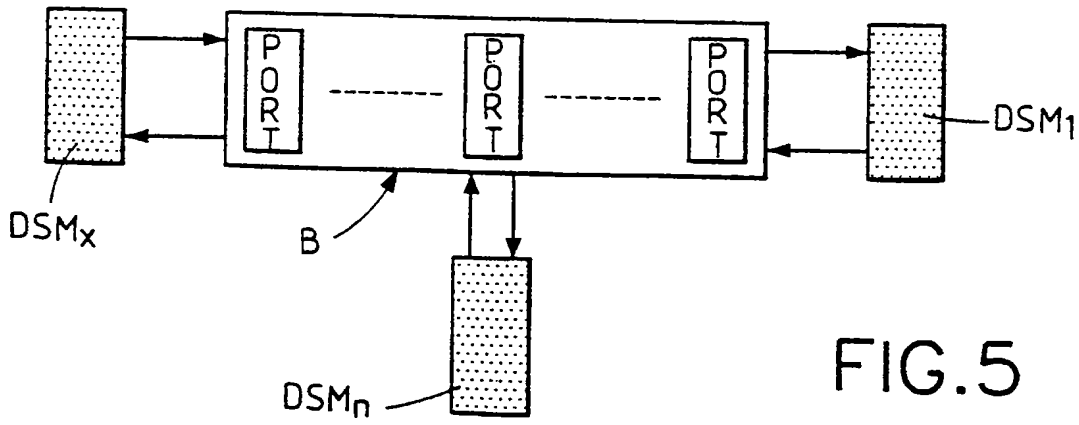


FIG.5

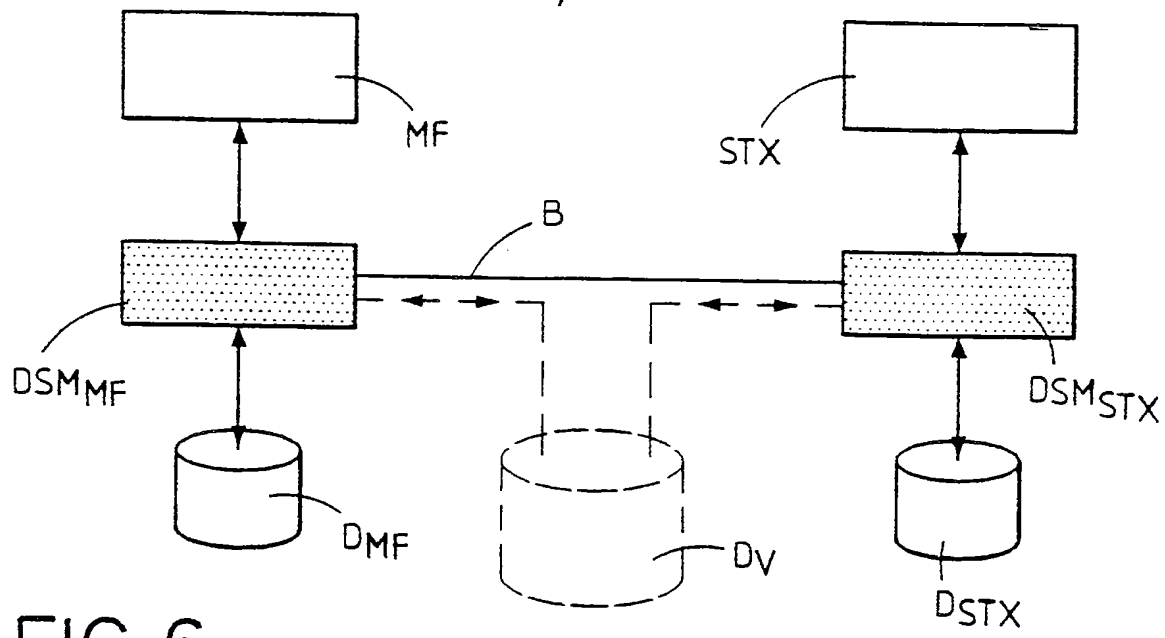


FIG.6

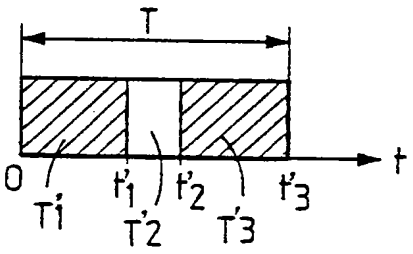


FIG.7

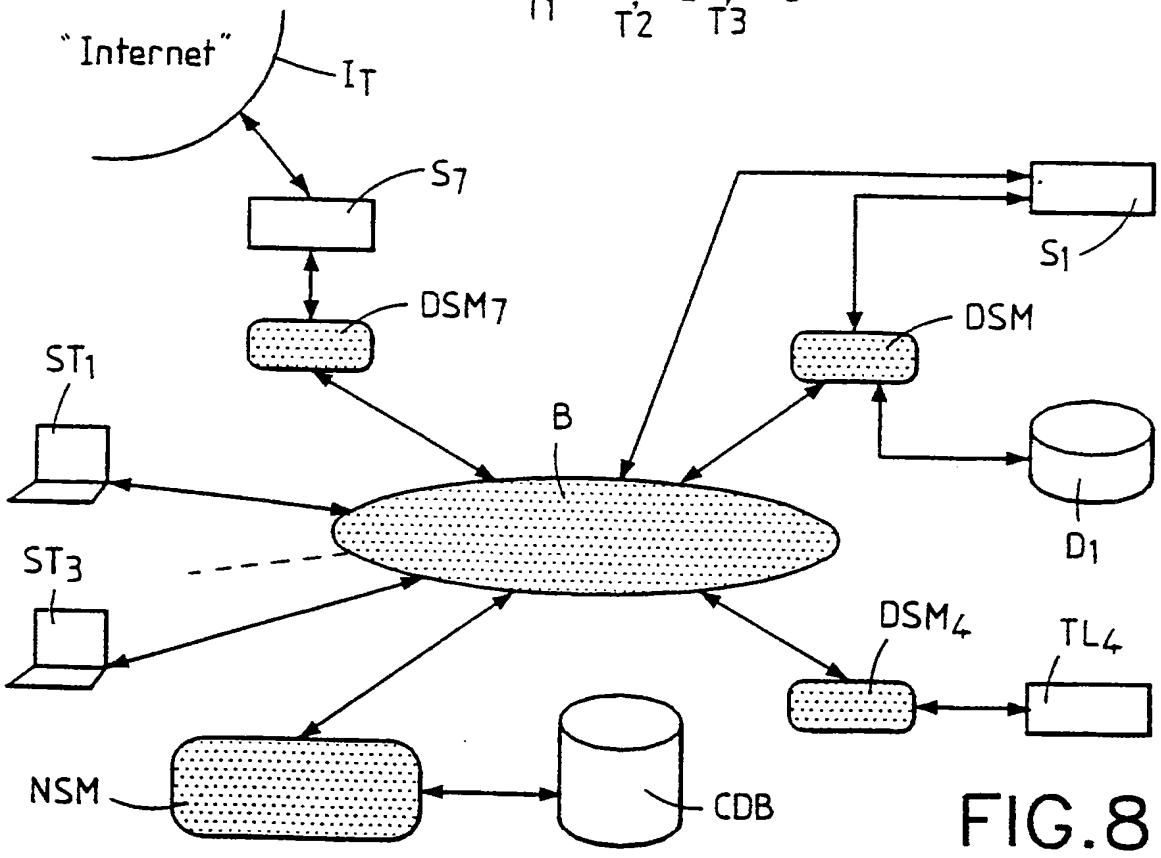


FIG.8

PCT ORGANISATION MONDIALE DE LA PROPRIETE INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets 6 : G06F 17/30 — A3

(11) Numéro de publication internationale: WO 98/33297

(43) Date de publication internationale: 30 juillet 1998 (30.07.98)

(21) Numéro de la demande internationale: PCT/FR98/00110

(22) Date de dépôt international: 22 janvier 1998 (22.01.98)

(30) Données relatives à la priorité:
97/00757 24 janvier 1997 (24.01.97) FR

(71) Déposant *(pour tous les Etats désignés sauf US)*: BULL S.A. [FR/FR]; 68, route de Versailles, F–78430 Louveciennes (FR).

(72) Inventeur; et
(75) Inventeur/Déposant *(US seulement)*: PEPING, Jacques [FR/FR]; 72, rue Victor Basch, F–78220 Viroflay (FR).

(74) Mandataire: DENIS, Hervé; Bull S.A., 68, route de Versailles, F–78430 Louveciennes (FR).

(81) Etats désignés: JP, US.

Publiée
Avec rapport de recherche internationale.
Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.

(88) Date de publication du rapport de recherche internationale:
17 septembre 1998 (17.09.98)

(54) Title: COMPUTER SYSTEM WITH DISTRIBUTED DATA STORING

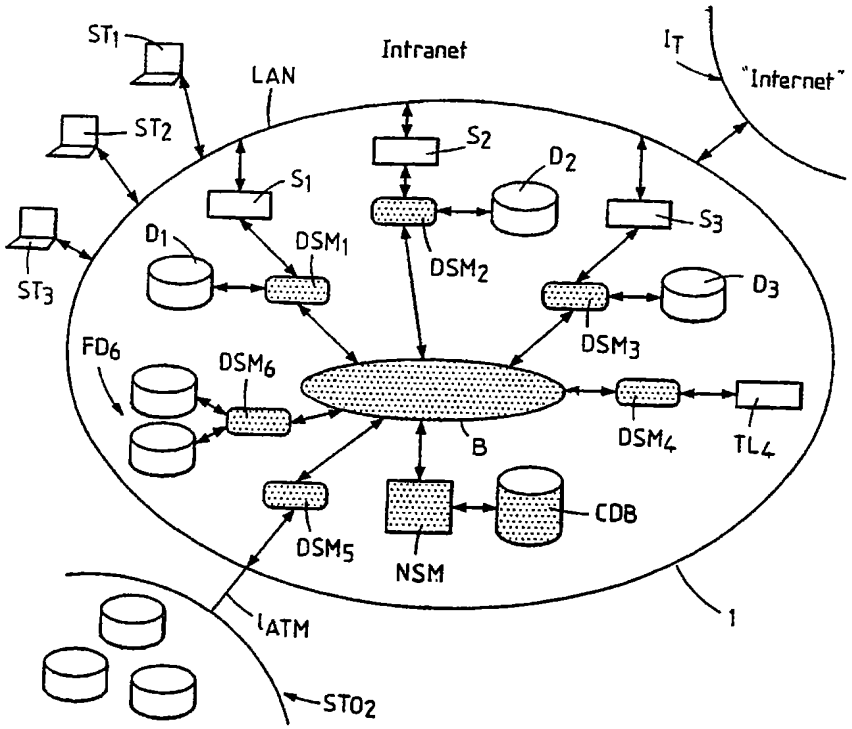
(54) Titre: SYSTEME INFORMATIQUE A STOCKAGE DE DONNEES DISTRIBUE

(57) Abstract

The invention concerns a computer system (1) in which each data storing resource (D_1 to D_3, FD_6, TL_4, STO_e) is under the control of a distributed management unit (DSM_1 to DSM_6), associated or not with a server (S_1 to S_3). The latter receives from a centralised management unit (NSM) control programmes dynamically allocating to it a virtual memory space comprising local storing units and all or part of the external storing resources. The invention is applicable to computer systems.

(57) Abrégé

Dans le système informatique (1) selon l'invention, chaque ressource de stockage de données (D_1 à D_3, FD_6, TL_4, STO_e) est sous la commande d'une unité d'administration distribuée (DSM_1 à DSM_6), associée ou non à un serveur (S_1 à S_3). Celle–ci reçoit d'une unité d'administration centralisée (NSM) des programmes de commande lui attribuant dynamiquement un espace mémoire virtuel comprenant des unités de stockage locales et tout ou partie des ressources de stockage externes.



UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	**ES**	Espagne	**LS**	Lesotho	**SI**	Slovénie
AM	Arménie	**FI**	Finlande	**LT**	Lituanie	**SK**	Slovaquie
AT	Autriche	**FR**	France	**LU**	Luxembourg	**SN**	Sénégal
AU	Australie	**GA**	Gabon	**LV**	Lettonie	**SZ**	Swaziland
AZ	Azerbaïdjan	**GB**	Royaume-Uni	**MC**	Monaco	**TD**	Tchad
BA	Bosnie-Herzégovine	**GE**	Géorgie	**MD**	République de Moldova	**TG**	Togo
BB	Barbade	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tadjikistan
BE	Belgique	**GN**	Guinée	**MK**	Ex-République yougoslave de Macédoine	**TM**	Turkménistan
BF	Burkina Faso	**GR**	Grèce			**TR**	Turquie
BG	Bulgarie	**HU**	Hongrie	**ML**	Mali	**TT**	Trinité-et-Tobago
BJ	Bénin	**IE**	Irlande	**MN**	Mongolie	**UA**	Ukraine
BR	Brésil	**IL**	Israël	**MR**	Mauritanie	**UG**	Ouganda
BY	Bélarus	**IS**	Islande	**MW**	Malawi	**US**	Etats-Unis d'Amérique
CA	Canada	**IT**	Italie	**MX**	Mexique	**UZ**	Ouzbékistan
CF	République centrafricaine	**JP**	Japon	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Pays-Bas	**YU**	Yougoslavie
CH	Suisse	**KG**	Kirghizistan	**NO**	Norvège	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	République populaire démocratique de Corée	**NZ**	Nouvelle-Zélande		
CM	Cameroun			**PL**	Pologne		
CN	Chine	**KR**	République de Corée	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Roumanie		
CZ	République tchèque	**LC**	Sainte-Lucie	**RU**	Fédération de Russie		
DE	Allemagne	**LI**	Liechtenstein	**SD**	Soudan		
DK	Danemark	**LK**	Sri Lanka	**SE**	Suède		
EE	Estonie	**LR**	Libéria	**SG**	Singapour		

INTERNATIONAL SEARCH REPORT

Interna al Application No
PCT/FR 98/00110

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 89 02631 A (DIGITAL EQUIPMENT CORP) 23 March 1989 see abstract see page 4, line 19 - page 5, line 7	1,2 7,16
A Y	EP 0 747 840 A (IBM) 11 December 1996 see abstract see column 2, line 10 - line 31 see column 5, line 13 - line 41 see column 6, line 59 - column 7, line 7	1-17 16
X Y	US 5 367 698 A (WEBBER NEIL F ET AL) 22 November 1994 see abstract see column 1 - column 3	1,2,15 7

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 July 1998	24/07/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Adkhis, F

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interna. al Application No
PCT/FR 98/00110

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EP 0 774 723 A (MATSUSHITA ELECTRIC IND CO LTD) 21 May 1997 see abstract see column 3, line 37 - column 4, line 12	1,2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern: al Application No
PCT/FR 98/00110

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 8902631 A	23-03-1989	CA 1312385 A	05-01-1993
		DE 3889904 D	07-07-1994
		DE 3889904 T	12-01-1995
		EP 0338041 A	25-10-1989
		JP 1502786 T	21-09-1989
		US 5408619 A	18-04-1995
EP 0747840 A	11-12-1996	US 5752246 A	12-05-1998
		CN 1143776 A	26-02-1997
		JP 9026973 A	28-01-1997
US 5367698 A	22-11-1994	NONE	
EP 0774723 A	21-05-1997	JP 10003421 A	06-01-1998

RAPPORT DE RECHERCHE INTERNATIONALE

Dema. Internationale No
PCT/FR 98/00110

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 6 G06F17/30

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultee (système de classification suivi des symboles de classement)
CIB 6 G06F

Documentation consultee autre que la documentationminimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de donnees électronique consultée au cours de la recherche internationale (nom de la base de donnees, et si cela est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Categorie	Identification des documents cités, avec, le cas échéant, l'indicationdes passages pertinents	no. des revendications visees
X Y	WO 89 02631 A (DIGITAL EQUIPMENT CORP) 23 mars 1989 voir abrégé voir page 4, ligne 19 - page 5, ligne 7	1,2 7,16
A Y	EP 0 747 840 A (IBM) 11 décembre 1996 voir abrégé voir colonne 2, ligne 10 - ligne 31 voir colonne 5, ligne 13 - ligne 41 voir colonne 6, ligne 59 - colonne 7, ligne 7	1-17 16
X Y	US 5 367 698 A (WEBBER NEIL F ET AL) 22 novembre 1994 voir abrégé voir colonne 1 - colonne 3	1,2,15 7

-/--

[X] Voir la suite du cadre C pour la finde la liste des documents

[X] Les documents de familles de brevets sont indiquésen annexe

' Catégories spéciales de documents cités:

"A" document définissant l'état général de latechnique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date dedépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendcation de priorité ou cité pour déterminer la date depublication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôtinternational, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après ladate de dépôt international ou la date de priorite et n'appartenenant pas à l'état de la technique pertinent, mais citépour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famillede brevets

Date à laquelle la recherche internationale a éteeffectivement achevée	Date d'expédition du présent rapport de recherche internationale
14 juillet 1998	24/07/1998

Nom et adresse postale de l'administrationchargée de la recherche internationale	Fonctionnaire autorisé
Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Adkhis, F

1

Formulaire PCT/ISA/210 (deuxième feuille) (juillet 1992)

RAPPORT DE RECHERCHE INTERNATIONALE

Demar. .nternationale No
PCT/FR 98/00110

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités. avec.le cas échéant. l'indicationdes passages pertinents	no. des revendications visées
P,X	EP 0 774 723 A (MATSUSHITA ELECTRIC IND CO LTD) 21 mai 1997 voir abrégé voir colonne 3, ligne 37 - colonne 4, ligne 12	1,2

1

Formulaire PCT/ISA/210 (suite de la deuxieme feuille) (juillet 1992)

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demai Internationale No
PCT/FR 98/00110

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 8902631 A	23-03-1989	CA 1312385 A	05-01-1993
		DE 3889904 D	07-07-1994
		DE 3889904 T	12-01-1995
		EP 0338041 A	25-10-1989
		JP 1502786 T	21-09-1989
		US 5408619 A	18-04-1995
EP 0747840 A	11-12-1996	US 5752246 A	12-05-1998
		CN 1143776 A	26-02-1997
		JP 9026973 A	28-01-1997
US 5367698 A	22-11-1994	AUCUN	
EP 0774723 A	21-05-1997	JP 10003421 A	06-01-1998